

# Iterative Shrinkage Approach to Restoration of Optical Imagery

Elad Shaked, *Student Member, IEEE* and Oleg Michailovich, *Member, IEEE*

## Abstract

The problem of reconstruction of digital images from their degraded measurements is regarded as a problem of central importance in various fields of engineering and imaging sciences. In such cases, the degradation is typically caused by the resolution limitations of an imaging device in use and/or by the destructive influence of measurement noise. Specifically, when the noise obeys a Poisson probability law, standard approaches to the problem of image reconstruction are based on using fixed-point algorithms which follow the methodology first proposed by Richardson and Lucy. The practice of using these methods, however, shows that their convergence properties tend to deteriorate at relatively high noise levels. Accordingly, in the present paper, a novel method for de-noising and/or de-blurring of digital images corrupted by Poisson noise is introduced. The proposed method is derived under the assumption that the image of interest can be sparsely represented in the domain of a linear transform. Consequently, a shrinkage-based iterative procedure is proposed, which guarantees the solution to converge to the global maximizer of an associated maximum-a-posteriori criterion. It is shown in a series of computer-simulated experiments that the proposed method outperforms a number of existing alternatives in terms of stability, precision, and computational efficiency.

## Index Terms

Deconvolution, sparse representations, iterative shrinkage, Poisson noise, and maximum-a-posteriori estimation.

Manuscript received Month XX, 2009. This research was supported by a Discovery grant from NSERC – The Natural Sciences and Engineering Research Council of Canada. Information on various NSERC activities and programs can be obtained from <http://www.nserc.ca>.

E. Shaked and O. Michailovich are with the School of Electrical and Computer Engineering, University of Waterloo, Canada N2L 3G1 (phone: 519-888-4567; e-mails: [eshaked](mailto:eshaked), [olegm@uwaterloo.ca](mailto:olegm@uwaterloo.ca)).

# Iterative Shrinkage Approach to Restoration of Optical Imagery

## I. INTRODUCTION

The image formation model of many key imaging modalities rely on the notion of *event counts*. The latter, for example, quantifies the number of gamma photons which pass through a single slit of the collimator of a gamma camera in positron emission tomography (PET) and single positron emission computer tomography (SPECT) [1]–[3]. In addition, statistical models of the same type are routinely used in optics to account for the process of “counting” the number of optical photons registered by a sensor of a (CCD) camera [4], [5]. Confocal microscopy [6], astronomical imaging [7], and turbulent imaging [8] are among other fields of imaging sciences where the notion of event counts is standardly used.

In the most general setup, the image formation model of the above-mentioned imaging modalities assumes that a data image is formed as a blurred version of the original (i.e. true) image contaminated by Poisson noise. Specifically, if  $g(\mathbf{x}) \in \mathbb{Z}_+$  (with  $\mathbf{x} \in \Omega \subseteq \mathbb{Z}^2$ ) denotes the observed Poisson counts image and  $f(\mathbf{x}) \in \mathbb{R}_+$  is its corresponding original counterpart, then the image formation model can be formally expressed as given by [9, Section 7.3]

$$g = \mathcal{P} \{ \mathbf{H}[f] \}, \quad (1)$$

where  $\mathcal{P}$  stands for the operation of contamination of  $\mathbf{H}[f]$  by Poisson noise and  $\mathbf{H}$  denotes the operator of convolution with a (known) point spread function (PSF)  $h$ , where  $h$  is assumed to be positive and mean preserving. Accordingly, to recover the original image  $f$  from  $g$ , the combined effect of  $\mathcal{P}\{\mathbf{H}[\cdot]\}$  has to be inverted.

The model of (1) has long been in use in a variety of different applications. For example, in [6] the model describes the formation of microscopic images, which are subsequently enhanced by means of a maximum-likelihood (ML) procedure, also known as the Richardson-Lucy algorithm [10], [11]. The same model and the same reconstruction approach has been used in the field of nuclear imaging, where they have been employed for recovery of PET and SPECT scans [12], [13]. It was argued in [14], however, that the ML approach may lead to either unstable or noisy results in the case of poorly conditioned operators  $\mathbf{H}$ . To alleviate this problem, it was proposed in [14] to replace the ML approach by maximum-a-posteriori (MAP) estimation, which provides means to regularize the inversion of  $\mathbf{H}$  via incorporating some prior information about  $f$  in the process of its reconstruction. Specifically, [14] suggested using either the Tikhonov-Miller or bounded variation models for  $f$ . Unfortunately, as will be shown in the experimental part of the present paper, the above algorithms cannot fully alleviate the instability problem intrinsic in the inversion of (1). To overcome this deficiency, the work in [15] suggested a different approach also based on the bounded variation model which leads to a minimum *total-variation* (TV) solution for the true image. In this case, to find a solution to (1), a variable splitting procedure of [16] was employed. Despite a substantial improvement in the stability of reconstruction, the computational cost of [15] is relatively high, which undermines

the applicability of this method for processing large data sets. Moreover, the algorithm of [15] may produce non-positives solutions, which are not natural in the case of Poisson imaging. To overcome the computational inefficiency of [15], a modification of the variable splitting method was proposed in [17]. Unfortunately, this modification seems to “restore” the instability concerns.

A different class of reconstruction methods has been recently proposed based on the assumption of *compressibility* of the true image  $f$ . This assumption suggests that  $f$  can be *sparsely represented* either in the spatial domain or in the domain of a certain linear transform. Thus, for example, [18] analyzes three different penalty functions, whose role is to impose the sparsity constraint on the estimated images. Although the proposed algorithms have exceptional stability and computation properties, they require the original image to be sparse in the spatial domain – the assumption which does not seem to be natural for a large spectrum of practical images. In this regard, a more general assumption has been used in [19], where the true images are supposed to be sparsely represented in the domain of a tight frame. This method, however, is based on “gaussianization” of the Poisson noise by means of a *variance stabilization transform* (VST) [20] that is known to perform inadequately in low-count settings. Another work that employed the assumption on  $f$  to be sparsely representable in the domain of an orthonormal transform is [21], where the sparsity-constrained Poisson inverse problem is solved in an approximate way by minimizing a sequence of  $\ell_2 - \ell_1$  objective functions. However, as acknowledged by the authors of [21], their algorithm has a drawback of slow convergence, which should be addressed in future research. As well, one should also mention the method of [22], in which an estimate of the true image  $f$  is obtained based on maximization of an ML criterion penalized by the  $\ell_0$ -norm of the platelet coefficients of the estimate. It is not clear, however, if the method of [22] could be used for a different type of image representations (rather than platelets), let alone the necessity to augment the performance of [22] by the computationally expensive cycle-spinning procedure of [23] to optimize the quality of resulting estimates.

In the case when the effect of  $\mathbf{H}$  can be neglected, a different family of reconstruction methods has been suggested based on the use of VST [20]. In particular, in such methods, the data image  $g$  is first preprocessed by a VST that transforms the Poisson noise into approximately homogeneous and Gaussian [24]–[27]. Subsequently, the true image is recovered from the “gaussianized” data using an appropriate statistical framework. Among other reconstruction techniques for solving (1) with  $\mathbf{H} \approx \mathbf{I}$  are wavelet Wiener filtering [28], hypothesis testing [29], as well as empirical Bayesian approaches [30], [31].

In the current paper, a different method for reconstruction of the true image  $f$  in (1) is introduced. As opposed to the algorithms described above, the proposed method is exceptional for it concurrently fulfills a number of essential objectives, *viz.*

- 1) *Exactness*: No auxiliary transformations and/or approximations are applied to the data image  $g$  to modify the properties of measurement noise, and therefore the method is applied under realistic statistical assumptions regarding the noise nature.
- 2) *Generality*: The image formation model employed by the proposed method is designed to accommodate a number of reconstruction scenarios, namely de-noising, de-blurring, and a combination thereof. As a result, the

very same procedure can be applied to recover an image of interest for a spectrum of different degradations.

- 3) *Versatility*: The proposed reconstruction is carried out under the assumption that the original image  $f$  can be sparsely represented in the domain of a linear transform. This assumption is currently recognized to be superior to many alternative models, including the Tikhonov-Miller and total variation models [32].
- 4) *Efficiency*: The proposed solution is based on the concept of *iterative shrinkage* [32]–[34] as a modern, computationally efficient, and stable procedure for solving a variety of inverse problems. In this way, the present contribution extends the theory of iterative shrinkage (aka thresholding) to the case of problems concerned with Poisson noises.
- 5) *Uniqueness*: Under a certain setup, the proposed algorithm is guaranteed to converge to the global maximizer of a maximum-a-posteriori criterion, thereby providing a unique solution to the reconstruction problem at hand.

Moreover, in the experimental part of the paper, it is shown that the proposed reconstruction method outperforms a number of alternative algorithms in terms of normalized mean-square error (NMSE), SSIM quality index [35], as well as its stability and computational efficiency.

The rest of the paper is organized as follows. Section II describes model assumptions and derives an optimization framework to be used in the following sections. Section III details a bound optimization approach, which results in an iterative shrinkage procedure. Section IV provides a proof of convergence of the proposed algorithm, while Section V presents a number of *in silico* reconstruction results. Section VI finalizes the paper with a discussion and conclusions.

## II. ESTIMATION FRAMEWORK

### A. Image modeling

The proposed approach for the reconstruction of  $f$  in (1) is based on the assumption that  $f$  can be sparsely represented in the domain of a certain linear transform. Specifically, let  $\{\varphi_k\}_{k \in \mathcal{I}}$  be a *frame* in the space of the original image  $f$  [36]. (Note that  $\mathcal{I}$  above denotes a set of frame indices, whose definition depends on a specific choice of  $\varphi_k$ .) Using the frame,  $f$  can be represented according to

$$f = \Phi(c) : f \mapsto \sum_{k \in \mathcal{I}} c_k \varphi_k, \quad (2)$$

where  $c \in \ell_2(\mathcal{I})$  denotes the representation coefficients of  $f$  in  $\text{span}\{\varphi_k\}_{k \in \mathcal{I}}$ , such that the number of *significant*  $c_k$  is much smaller than  $\#\mathcal{I}$ , i.e. the total number of  $\varphi_k$ . It is worthwhile noting that the model of (2) is standard in the theory of sparse representation which has firmly reserved a leading position among the modern tools of signal and image processing. The value of sparse representations has been demonstrated in numerous fields and applications, which include signal modeling [37], compressed sensing [38], [39], independent component analysis and blind source separation [40], [41], inverse problems [33], [42], signal and image de-noising [43], [44], morphological component analysis [45] as well as its earlier version in [46], just to name a few.

When expressed as a function of the representation coefficients  $c$ , the image formation model (1) becomes

$$g = \mathcal{P} \{ \mathbf{H}[\Phi(c)] \} = \mathcal{P} \{ \mathbf{A}[c] \}, \quad (3)$$

where  $\mathbf{A} = \mathbf{H} \cdot \Phi$  is a composition map from  $\ell_2(\mathcal{I})$  to the signal space, which represents the combined effect of image synthesis and blur. In the discussion below, we will be mainly interested in the following three settings:

- *Setting 1:* The blur is negligible and hence  $\mathbf{A} \equiv \Phi$ . In this case, the problem becomes that of *image de-noising*.
- *Setting 2:* The basis  $\Phi$  is the canonical (Dirac) basis (i.e.  $\Phi = \mathbf{I}$ ) and hence  $\mathbf{A} \equiv \mathbf{H}$ . In this case, the image  $f(\mathbf{x})$  is identified with its representation coefficients  $c$ , and the problem becomes that of *sparse deconvolution*.
- *Setting 3:* Neither  $\mathbf{H}$  nor  $\Phi$  can be simplified/neglected. In this case, the estimation problem at hand becomes that of *sparse reconstruction*.

For the sake of generality, in what follows the specific nature of  $\mathbf{A}$  will not be specified until the experimental part of the paper, where different reconstruction examples are presented. In particular,  $\mathbf{A}$  is considered to be a linear map from the discrete domain of the representation coefficients  $c \in \ell_2(\mathcal{I})$  to the space of blurred images.

It is well known that in the case of a poorly conditioned operator  $\mathbf{A}$ , the problem of recovering  $c \in \ell_2(\mathcal{I})$  from noisy measurements of  $\mathbf{A}[c]$  can be highly unstable (in the sense that there will not be a continuous dependency between the data and an estimate of  $c$ ). In order to overcome this deficiency, we employ the framework of MAP estimation which provides the most likely solution, given the observed data and a reasonable assumption regarding the statistical nature of the true coefficients  $c$  [47]. The details on this approach are presented next.

### B. MAP estimation

Since most of the modern imaging devices work with finite data, it seems to be appropriate to restrict the subset  $\Omega \subset \mathbb{Z}^2$  to be a finite lattice. Accordingly, the measurement  $g$  and the original image  $f$  are assumed to be of size  $N \times M$  and  $K \times L$ , respectively. In this case, the range of  $\mathbf{A}$  becomes a set of bounded  $N \times M$  matrices.

To compensate for the information loss inflicted by the poor conditioning of  $\mathbf{A}$ , the MAP estimator takes advantage of *a priori* knowledge regarding the quantity of interest, i.e.  $c$ . In the case at hand, such knowledge is specified in the form of a prior probability distribution  $P(c)$ , which leads to the following definition of the MAP estimator

$$c_{MAP} = \underset{c}{\operatorname{argmax}} P(c | g) = \underset{c}{\operatorname{argmax}} P(g | c) P(c), \quad (4)$$

where  $P(g | c)$  represents the likelihood function pertaining to the observation model (1). In the case of Poisson noises, the likelihood function of a single measurement  $g(\mathbf{x}_i)$  at pixel  $\mathbf{x}_i$  is given by

$$P(g(\mathbf{x}_i) | c) = \frac{e^{-(\mathbf{A}[c])_i} (\mathbf{A}[c])_i^{g(\mathbf{x}_i)}}}{g(\mathbf{x}_i)!}, \quad (5)$$

where  $(\mathbf{A}[c])_i$  denotes the  $\mathbf{x}_i$ -th coordinate of  $\mathbf{A}[c]$  and  $g(\mathbf{x}_i) \in \mathbb{Z}_+$  is interpreted as the “number of counts” (e.g. the number of gamma photons registered by a gamma camera in nuclear imaging). Assuming that the values of  $g(\mathbf{x}_i)$  are independent and identically distributed (i.i.d.), the joint probability of the observed image  $g$  is given by

$$P(g | c) = \prod_{\mathbf{x}_i=1}^{NM} \frac{e^{-(\mathbf{A}[c])_i} (\mathbf{A}[c])_i^{g(\mathbf{x}_i)}}}{g(\mathbf{x}_i)!} \quad (6)$$

To complete the model, the coefficients of  $c = \{c_k\}_{k \in \mathcal{I}}$  are assumed to be identically distributed according to a Generalized Gaussian (GG) probability law [48], so that the joint probability of  $c$  can be defined as

$$P(c) = \prod_{k \in \mathcal{I}} \frac{p}{2\beta \Gamma(p^{-1})} e^{-(|c_k|/\beta)^p} \quad (7)$$

While  $\beta > 0$  controls the variance of  $c_k$ , the value of  $0 < p < \infty$  determines the appearance of  $c$  in terms its sparsity (with smaller values of  $p$  resulting in “heavier” tails of the corresponding probability density function (pdf), thereby allowing larger values of  $c_k$  to be occasionally drawn). In particular, the choice of  $p = 1$  results in a Laplacian distribution, which is commonly used to describe the behavior of sparse representation coefficients [37], [43]. Subsequently, the final expression for the MAP estimate becomes

$$c_{MAP} = \underset{c}{\operatorname{argmax}} \prod_{\mathbf{x}_i=1}^{NM} \left\{ \frac{e^{-(\mathbf{A}[c])_i} (\mathbf{A}[c])_i^{g(\mathbf{x}_i)}}}{g(\mathbf{x}_i)!} \right\} \prod_{k \in \mathcal{I}} \left\{ e^{-(|c_k|/\beta)^p} \right\}. \quad (8)$$

It is conventional to convert the maximization problem (8) into a minimization problem through applying the log-transform to the posterior probability in (8), followed by inverting the sign of the expression thus obtained. In this case, the MAP estimate can be re-expressed as given by

$$c_{MAP} = \underset{c}{\operatorname{argmin}} \{E(c)\},$$

$$E(c) = \langle \mathbf{1}, \mathbf{A}[c] \rangle - \langle g, \log(\mathbf{A}[c]) \rangle + \gamma \|c\|_p^p, \quad \gamma \triangleq 1/\beta^p, \quad (9)$$

where  $\mathbf{1}$  stands for an  $N \times M$  matrix of ones,  $\langle \cdot, \cdot \rangle$  stands for the standard inner product in  $\mathbb{R}^{N \times M}$  and  $\|c\|_p^p = \sum_{k \in \mathcal{I}} |c_k|^p$  is the  $\ell_p$ -norm of the representation coefficients.

It should be noted that the likelihood model of (6) interprets the true image value  $f(\mathbf{x}_i)$  as the mean value of the corresponding random observation  $g(\mathbf{x}_i)$ . Moreover, since in the case of the Poisson distribution, the first and the second moments of the distribution are equal, the values  $f(\mathbf{x}_i)$  should be assumed to be nonnegative. The latter assumption restricts the domain of  $E(c)$  in (9) to be now defined as

$$\mathbf{dom} E = \{c \in \ell_2(\mathcal{I}) \mid \Phi[c] \geq 0\}. \quad (10)$$

Note that  $E(c)$  is convex over  $\mathbf{dom} E$ . Moreover, it is strictly convex over  $\mathbf{dom} E$  if  $p > 1$  or  $g \succ 0$  and  $\mathbf{A}$  is non-degenerate (albeit, possibly, ill-conditioned). In the latter case, the functional  $E(c)$  in (9) admits a unique minimizer (note that the set  $\mathbf{dom} E$  is convex) [49, Ch. 11], which can be found by any algorithm which is guaranteed to converge to a stationary point of  $E(c)$ . On the other hand, when either  $g$  is not strictly positive or  $\mathbf{A}$  possesses a non-trivial null-space, the convexity of  $E(c)$  is not strict. In such a situation, the uniqueness of a minimizer of  $E(c)$  for  $p = 1$  (which is the case of primary concern in this paper) can be re-established by replacing the  $\ell_1$ -norm in (9) by its smooth relaxation, e.g., by  $\langle \mathbf{1}, \sqrt{c^2 + \epsilon} \rangle$ , for some  $\epsilon \ll 1$  [42]. It should be noted, however, that as long as practical aspect of the image reconstruction is concerned, it has been observed that, for the case of the method proposed in this paper, the above relaxation seems to be unnecessary even in the case of degenerate  $\mathbf{H}$  and  $g \succeq 0$ .

It should be noted that the problem of minimizing  $E(c)$  over the set defined by (10) does not necessarily have to be formulated as a constrained minimization problem. This is because, for nonnegative data images, the term  $-\langle g, \log(\mathbf{A}[c]) \rangle$  in (9) works similar to a log-barrier function which forces the solution to stay within the feasible region defined by the condition  $\Phi[c] \succeq 0$ . In fact, strong theoretical guarantees for the existence and attainability of a unique minimizer of  $E(c)$  follow directly from the theory of interior-point methods [49, Ch.11].

A possible solution to computing  $c_{MAP}$  could be by means of standard optimization techniques, such as gradient-based methods [50]. It was recently argued, however, that such general purpose tools might be ineffective when applied to the problem at hand, since they do not exploit properly the sparse structure of the desired solution [32]. In the following section, a different approach to finding  $c_{MAP}$  is detailed based on the methodology of iterative shrinkage [32]–[34], [42].

### III. BOUND OPTIMIZATION AND ITERATIVE SHRINKAGE

The non-differentiability of the  $\ell_p$ -norm for the case  $p = 1$  (which forms the main focus of this paper) makes it impossible to solve (9) by means of gradient-based optimization methods. Fortunately, an effective solution to the problem can be derived using the majorization-minimization (MM) approach of [51] (also known as the method of bound optimization [52]). The first step in applying this method consists in replacing the original problem of minimizing  $E(c)$  by a problem of minimizing a *surrogate* functional  $Q(c, c_0)$  (with  $c_0 \in \ell_2(\mathcal{I})$  being an arbitrary set of coefficients). Specifically, let the new functional  $Q(c, c_0)$  have the following form

$$Q(c, c_0) = E(c) + \Psi(c, c_0), \quad (11)$$

where

$$\Psi(c, c_0) = \langle g, \log(\mathbf{A}[c]/\mathbf{A}[c_0]) \rangle - \langle \mathbf{A}^* [g/\mathbf{A}[c_0]], c - c_0 \rangle_{\ell_2(\mathcal{I})} + \frac{\mu}{2} \|c - c_0\|_2^2, \quad (12)$$

with the “slash” to be interpreted as an element-wise division,  $\mathbf{A}^*$  being the adjoint of  $\mathbf{A}$ , and  $\langle \cdot, \cdot \rangle_{\ell_2(\mathcal{I})}$  standing for the inner product in  $\ell_2(\mathcal{I})$ . It goes without saying that the surrogate functional in (12) has to obey a number of critical constraints which will be detailed in Section IV.

To minimize  $Q(c, c_0)$  with respect to  $c$ , its gradient is computed first according to

$$\begin{aligned} \nabla Q(c, c_0) = & \mathbf{A}^* [1] - \mathbf{A}^* [g/\mathbf{A}[c]] + p \gamma |c|^{p-1} \text{sign}(c) + \\ & + \mathbf{A}^* [g/\mathbf{A}[c]] - \mathbf{A}^* [g/\mathbf{A}[c_0]] + \mu(c - c_0). \end{aligned} \quad (13)$$

Subsequently, canceling the similar terms with opposite signs and equating the gradient to zero yields

$$c + \frac{p\gamma}{\mu} |c|^{p-1} \text{sign}(c) = c_0 + \frac{1}{\mu} \mathbf{A}^* [g/\mathbf{A}[c_0] - 1]. \quad (14)$$

It should be noted that the above analysis is relevant to the differentiable case of  $p > 1$ . For the case of  $p = 1$ , the first order optimality condition (14) has to be replaced by its more general version requiring

$$c + \frac{\gamma}{\mu} \partial \|c\|_1 \ni c_0 + \frac{1}{\mu} \mathbf{A}^* [g/\mathbf{A}[c_0] - 1], \quad (15)$$

where  $\partial\|c\|_1$  denotes the subdifferential of  $\|c\|_1$ . Remarkably, both (14) and (15) have the same closed form solution. Particularly, given the *inverse*  $\mathcal{S}_{p,\gamma,\mu}(c)$  of the function on the left-hand side of (14), the optimal  $c$  that minimizes  $Q(c, c_0)$  is given by

$$c = \mathcal{S}_{p,\gamma,\mu} \left( c_0 + \frac{1}{\mu} \mathbf{A}^* \left[ \frac{g - \mathbf{A}[c_0]}{\mathbf{A}[c_0]} \right] \right). \quad (16)$$

The functional in (16) is analogous to the functional used in the existing iterative shrinkage methods, which have been derived under the assumption of Gaussian noises [32]–[34], [42]. Just like in the latter studies, we propose to find a minimizer of the original functional in (9) iteratively, *viz.*

$$c_{t+1} = \mathcal{S}_{p,\gamma,\mu} \left( c_t + \frac{1}{\mu} \mathbf{A}^* \left[ \frac{g - \mathbf{A}[c_t]}{\mathbf{A}[c_t]} \right] \right), \quad (17)$$

where  $t$  stands for the iteration index. The convergence of the above iteration scheme to a global minimizer of (9) is proven in the next section. Before turning to the proof, we note that setting  $p = 1$  seems to be a reasonable choice in practice, since it guarantees the convexity of  $E(c)$  on one hand, and leads to a sparse solution on the other. It is worthwhile noting that, in this case, the function  $\mathcal{S}_{p,\gamma,\mu}$  becomes the notorious *soft thresholding* rule [43], which is defined as

$$\mathcal{S}_{1,\gamma,\mu}(c) = \begin{cases} (|c| - \gamma/\mu) \text{sign}(c), & \text{if } |c| \geq \gamma/\mu \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

Finally, it is worthwhile noting that the computational cost in (17) is mainly determined by the cost of applying the composition transform  $\mathbf{A} = \mathbf{H} \cdot \Phi$  and its adjoint. While the convolution operator can be computed efficiently, e.g., by means of the fast Fourier transform [53], the application of  $\Phi$  and  $\Phi^*$  depends on the type of transformation in use. Fortunately, most of the relevant transforms (such as wavelet [54], ridgelet [55], and curvelet [39] transforms) admit computationally efficient implementations, which are typically of a logarithmic complexity at most. A formal comparison between the computational efficiencies of the proposed and reference methods is given in Section V.

#### IV. CONVERGENCE ANALYSIS

To demonstrate that the proposed algorithm constitutes a viable alternative to traditional approaches, its convergence properties need to be analyzed next. As was shown in [51], the convergence can be guaranteed provided that the surrogate functional  $Q(c, c_t)$  satisfies the following conditions:

$$\begin{aligned} 1) & \quad Q(c, c_t) \geq E(c), \quad \forall c \in \mathbf{dom} E \\ 2) & \quad Q(c_t, c_t) = E(c_t) \end{aligned} \quad (19)$$

with  $c_t$  being an arbitrary but fixed reference point in  $\mathbf{dom} E$ .

Before verifying the above conditions, it is instructive to show first that minimizing  $Q(c, c_t)$  with respect to  $c$  yields a reduction in the value of the original functional  $E(c)$ . To this end, let  $c_{t+1}$  denote a minimizer of  $Q(c, c_t)$ , i.e.  $c_{t+1} = \text{argmin}_c Q(c, c_t)$ . For such a  $c_{t+1}$  and  $Q(c, c_t)$  obeying (19), one then obtains

$$E(c_{t+1}) = Q(c_{t+1}, c_t) + E(c_{t+1}) - Q(c_{t+1}, c_t) \leq Q(c_t, c_t) + E(c_t) - Q(c_t, c_t) = E(c_t), \quad (20)$$



resulting in

$$E(c_{t+1}) \leq E(c_t). \quad (21)$$

Thus, the bound-optimization algorithm (17) guarantees a reduction in the value of  $E(c_t)$  at each iteration provided  $Q(c, c_t)$  obeys the conditions in (19). It should be noted, however, that for the value of  $E(c_{t+1})$  to decrease, it is sufficient to require that  $Q(c, c_t) \geq E(c)$  *only* at  $c = c_{t+1}$ . The latter seems to be much less restrictive condition as compared to Condition 1 in (19). (We will take advantage of this fact later in this section.)

The fact that Condition 2 in (19) is satisfied by the proposed surrogate functional  $Q(c, c_t)$  can be verified by direct substitutions. Moreover, it is important to point out that the function  $\Psi$  in (11) obeys

$$\Psi(c_t, c_t) = 0 \quad \text{and} \quad \nabla \Psi(c, c_t) \big|_{c=c_t} = 0, \quad (22)$$

which suggests that  $\Psi(c, c_t)$  has an extremum at  $c = c_t$  that is equal to zero. Consequently, if  $\Psi(c, c_t)$  is a convex function, Condition 1 will be automatically fulfilled. The convexity, on the other hand, can be determined from the properties of the Hessian operator corresponding to  $\Psi(c, c_t)$ , which can be shown to be equal to

$$\nabla^2 \Psi(c) = \mu \mathbf{I} - \mathbf{A}^* \text{diag} \left( \frac{g}{(\mathbf{A}[c])^2} \right) \mathbf{A}. \quad (23)$$

Note that, in the expression above,  $\mathbf{I}$  stands for the identity operator, while the diagonal operator  $\text{diag} (g/(\mathbf{A}[c])^2)$  is defined in the standard way as

$$\text{diag} \left( \frac{g}{(\mathbf{A}[c])^2} \right) [y] \equiv \frac{g}{(\mathbf{A}[c])^2} \cdot y, \quad (24)$$

for any arbitrary  $y \in \mathbb{R}^{N \times M}$ . Provided the elements of  $g/(\mathbf{A}[c])^2$  are bounded and  $\mathbf{A}$  is compact, the Hessian in (23) is necessarily a self-adjoint and compact operator. Moreover, the minimal eigenvalue  $\lambda_{min}$  of the Hessian can be shown to be bounded by

$$\lambda_{min} \geq \mu - \|g/(\mathbf{A}[c])^2\|_{\infty} \lambda_{max}(\mathbf{A}^* \mathbf{A}), \quad (25)$$

where  $\|\cdot\|_{\infty}$  stands for the supremum norm and  $\lambda_{max}(\mathbf{A}^* \mathbf{A})$  denotes the maximum eigenvalue of  $\mathbf{A}^* \mathbf{A}$ . Thus, as long as

$$\mu \geq \|g/(\mathbf{A}[c])^2\|_{\infty} \lambda_{max}(\mathbf{A}^* \mathbf{A}), \quad (26)$$

the Hessian  $\nabla^2 \Psi(c)$  is positive definite, in which case  $\Psi(c, c_t)$  is convex, and hence Condition 1 of (19) is satisfied. It should be pointed that the finiteness of  $\mu$  is always guaranteed by the fact that  $c \in \text{dom } E$ . The value of  $\mu$  in (26), however, is defined as a function of  $c$ , and therefore it would be quite problematic (if possible at all) to determine  $\mu$  from (26), had we decided to do so. In this sense, the condition (26) should be regarded as merely a “proof of existence”.

It turns out that in practical cases there are much simpler means to determine a value of  $\mu$  that guarantees a reduction in the value of  $E(c)$  with respect to  $E(c_t)$ . First, we note that according to (20), it is sufficient to find a

$\mu_t$  that results in<sup>1</sup>

$$\frac{\mu_t}{2} \|c_{t+1} - c_t\|_2^2 \geq \langle \mathbf{A}^* [g/\mathbf{A}[c_t]], c_{t+1} - c_t \rangle_{\ell_2(\mathcal{I})} - \langle g, \log(\mathbf{A}[c_{t+1}]/\mathbf{A}[c_t]) \rangle, \quad (27)$$

thereby guaranteeing that  $\Psi(c_{t+1}, c_t) \geq 0$ , and therefore  $E(c_{t+1}) \leq E(c_t)$ . This implies the following practical way to find an acceptable  $\mu_t$ . Let the right-hand side of (27) be denoted by  $F(c_{t+1}, c_t)$ , which is a computable quantity provided the values of  $c_t$  and  $c_{t+1}$ . Consequently, a suitable value of  $\mu_t$  can be found using the following algorithm.

---

**Algorithm 1** Finding a suitable scaling parameter  $\mu_t$

---

```

1: Preset:  $\nu = 1, 0 < \alpha < 1$ 
2: Compute:  $c_{t+1} = \mathcal{S}_{1,\gamma,1}(c_t + \mathbf{A}^* [g/\mathbf{A}[c_t] - 1])$ 
3: if  $\|c_t - c_{t+1}\|_2^2 \geq 2 F(c_{t+1}, c_t)$  then
4:   while  $\nu \|c_t - c_{t+1}\|_2^2 \geq 2 F(c_{t+1}, c_t)$  do
5:      $\nu \leftarrow \alpha \nu$ 
6:      $c_{t+1} = \mathcal{S}_{1,\gamma,\nu}(c_t + \frac{1}{\nu} \mathbf{A}^* [g/\mathbf{A}[c_t] - 1])$ 
7:   end while
8:    $\mu_t = \nu/\alpha$ 
9: else
10:  while  $\nu \|c_t - c_{t+1}\|_2^2 < 2 F(c_{t+1}, c_t)$  do
11:     $\nu \leftarrow \nu/\alpha$ 
12:     $c_{t+1} = \mathcal{S}_{1,\gamma,\nu}(c_t + \frac{1}{\nu} \mathbf{A}^* [g/\mathbf{A}[c_t] - 1])$ 
13:  end while
14:   $\mu_t = \nu$ 
15: end if
16: return  $\mu_t$ 

```

---

Algorithm 1 has been designed to find a *minimal possible*  $\mu$  (with the accuracy of  $\log \alpha$  in the logarithmic scale) that guarantees that<sup>2</sup>  $E(c_{t+1}) \leq E(c_t)$ . It should be emphasized that the proof of existence in (26) assures that an acceptable value of  $\mu_t$  can always be found. The form of the shrinkage operator in (17) suggests that smaller values of  $\mu_t$  result in more substantial shrinkage, which in turn leads to more sizable changes in  $c_{t+1}$  with respect to  $c_t$ . The idea of maximizing the effect of shrinkage through minimizing the value of  $\mu_t$  lies in the heart of the method proposed in [56]. In the present case, the minimality of  $\mu_t$  is guaranteed by the design of Algorithm 1, whose only downside is in the extra calculations required. In practical scenarios, however, when one is working with a pool

<sup>1</sup>The subscript  $t$  is added intentionally in  $\mu_t$  to indicate that its value can be iteration-dependent.

<sup>2</sup>The accuracy can be improved via choosing  $\beta$  to be close to 1, which has a drawback of slowing down the convergence.

of images of similar nature, Algorithm 1 can be excluded by predefining the value of  $\mu$ . This can be done, for example, by applying Algorithm 1 to several images and determining a fixed value of  $\mu$  that satisfies (27).

## V. RESULTS

### A. Reference methods

In the experimental part of this paper, the proposed method is referred to as *Poisson Iterative Shrinkage* (PIS). As was argued earlier, PIS is conceptually parallel to the iterative shrinkage method developed in [33], [42] under the assumption of Gaussian noises, in which case the iterative shrinkage (referred to below as *Gaussian Iterative Shrinkage* (GIS)) has the form of

$$c_{t+1} = \mathcal{S}_{p,\gamma,\mu} \left( c_t + \frac{1}{\mu} \mathbf{A}^* [g - \mathbf{A}[c_t]] \right) \quad (28)$$

with  $\gamma = \sigma^2/\beta$  where  $\beta$  is the bandwidth parameter of the GG pdf in (7) and  $\sigma^2$  is the variance of the i.i.d. Gaussian noise contaminating the (blurred) measurements of the true image. The GIS algorithm is known to converge to the global minimizer of a MAP criterion provided  $\mu > \|\mathbf{A}^* \mathbf{A}\|$  [33].

Another reference method used in the comparison is the *Richardson-Lucy* (RL) algorithm, which is represented by the following iterative scheme

$$f_{t+1} = f_t \mathbf{H}^* \left[ \frac{g}{\mathbf{H}[f_t]} \right]. \quad (29)$$

Note that, as opposed to the GIS and PIS methods, the iterations in RL are performed on an estimated image  $f_t$  rather on the coefficients of its representation in a basis/frame. The iteration procedure in (29) is derived from a maximum-likelihood (ML) model under the assumption of Poisson noises. ML estimators, however, are known to result in degraded performance in the case of poorly conditioned  $\mathbf{H}$ . To alleviate this deficiency, [14] proposed to use the MAP framework to regularize the convergence in (29). This method assumes the true  $f$  to be in the space of bounded-variation images and, thus, the resulting iterative scheme minimizes the total variation (TV) norm of the estimate of  $f$ . Using the methodology of RL, the above minimization can be performed via

$$f_{t+1} = \frac{f_t}{1 - \gamma \operatorname{div} \left( \frac{\nabla f_t}{\|\nabla f_t\|} \right)} \mathbf{H}^* \left[ \frac{g}{\mathbf{H}[f_t]} \right], \quad (30)$$

where  $\gamma$  is a regularization parameter, which is to obey  $1 - \gamma \operatorname{div} \left( \frac{\nabla f_t}{\|\nabla f_t\|} \right) \succ 0$  to guarantee the reconstruction to be non-negative. Below, the algorithm of (30) is referred to as the Richardson-Lucy total variation (RLTV) method.

A different reference method that also takes advantage the TV regularization is [17]. This method is based on minimizing the same objective function as RLTV, subject to a non-negativity constraint on  $f$ . The solution in [17] uses the variable splitting technique of [16], which allows reducing the minimization problem to a few simpler subproblems. In what follows, the method is referred to as PIDSplit+ (which stands for Poisson image deconvolution by variable splitting with a positivity constraint).

Another reference approach used in the present study is that of [19]. Similarly to the method proposed in the present paper, [19] assumes  $f$  to be sparsely representable in the domain of a linear transform. Subsequently, the

algorithm is initialized by applying a VST (namely, the Anscombe transform) to the data image  $g$ , followed by recovering  $f$  as a solution to a standard  $\ell_2 - \ell_1$  minimization problem. In the present study, the above method has been implemented using a publicly available code (see <http://www.greyc.ensicaen.fr/~fdupe/>). Due to the fact that [19] incorporates a VST with the theory sparse representations, in the discussion that follows, this method is referred to as VSTSR.

The last reference method used in our comparative study is the one described in [21]. The method employs the sparsity assumption in the domain of an orthogonal wavelet transform. Subsequently, [21] solves the minimization problem of (9) by minimizing a sequence of quadratic approximations to a log penalty function. Following [21], the above method is referred to as sparse Poisson intensity reconstruction algorithm (SPIRAL).

Since RL, RLTV and PIDSplit+ tend to become unstable in the case of poorly conditioned  $\mathbf{H}$ , a common practice is to terminate their execution after a predefined number of iterations. In our experiments, their termination was performed at the point where the *normalized mean-squared error* (NMSE) (defined below) reached its minimum value. Needless to say that such termination is only possible under the conditions of controlled simulation studies, when the original images are known. In practical scenarios, however, the NMSE-optimal termination is generally impossible, which suggests that real-life reconstructions obtained with RL, RLTV and PIDSplit+ may be actually worse than the reconstructions demonstrated in the present paper. The proposed PIS algorithm, on the other hand, remains stable in the course of its convergence, which makes it possible to terminate the algorithm simply after a relative change in the value of  $E(c)$  drops below a predefined threshold (e.g.,  $10^{-6}$ ).

It is worthwhile noting that the reference methods above have been derived using different statistical approaches and assumptions. The motivation behind choosing these methods has been to compare the proposed method to the approaches based on the Poisson model (i.e. RL, RLTV, PIDSplit+, VSTSR, SPIRAL) as well as to those exploiting the idea of iterative shrinkage (i.e. GIS). The main properties of all the reconstruction methods under consideration are summarized in Table I.

TABLE I  
PROPERTIES OF THE RECONSTRUCTION METHODS UNDER COMPARISON

	Statistical framework	Prior model	Noise model
<b>RL</b>	ML	N/A	Poisson
<b>RLTV</b>	MAP	BV image $f$	Poisson
<b>PIDSplit+</b>	MAP	BV image $f$	Poisson
<b>VSTSR</b>	MAP	GG coefficients $c$	Poisson
<b>SPIRAL</b>	MAP	Laplacian coefficients $c$	Poisson
<b>GIS</b>	MAP	GG coefficients $c$	Gaussian
<b>PIS</b>	MAP	GG coefficients $c$	Poisson

TABLE II  
COMPUTATIONAL COMPLEXITY OF THE RECONSTRUCTION METHODS UNDER COMPARISON

	Computational operations per iteration
<b>PIS</b>	$2C(n) + 2R(n) + \mathcal{O}(n)$
<b>GIS</b>	$2C(n) + 2R(n) + \mathcal{O}(n)$
<b>RL</b>	$2C(n) + \mathcal{O}(n)$
<b>RLTV</b>	$2C(n) + \mathcal{O}(n)$
<b>PIDSplit+</b>	$4C(n) + \mathcal{O}(n)$
<b>VSTSR</b>	$2C(n) + \mathcal{O}(R(n))$
<b>SPIRAL</b>	$2C(n) + \mathcal{O}(n)$

### B. Computational complexity

The most computationally demanding operations used in the proposed and reference methods are associated with the convolution  $\mathbf{H}$  and frame  $\Phi$  operators. The convolution operator  $\mathbf{H}$  and its adjoint can be efficiently implemented by using, e.g., the FFT algorithm, which requires  $n \log n$  MAC operations, with  $n = NM$  representing the total number of pixels. Since the implementation of convolution and its related complexity depend on  $\mathbf{H}$  itself, let  $C(n)$  represent the total number of MAC operations required by computing  $\mathbf{H}$  and  $\mathbf{H}^*$ . In a similar manner, let  $R(n)$  represent the total number of MAC operations required by applying the frame operator  $\Phi$  and its adjoint. (Thus, for example,  $R(n) \sim \mathcal{O}(n)$  in the case of  $\Phi^*$  being an orthogonal wavelet transform). Consistent with the notations above, Table II summarizes the computational complexities required by *one iteration* of the proposed and reference methods. Needless to say, the overall complexity of the above methods will depend on the number of iterations required till their convergence. These numbers will be provided below for specific examples of image reconstruction.

### C. Comparison measures

The algorithms listed in Table I have been compared in terms of the NMSE defined as follows. Let  $f$  be a true image and  $\tilde{f}$  be an estimate of  $f$ . Then, the NMSE can be defined as

$$\text{NMSE} = \mathcal{E} \left\{ \frac{\|f - \tilde{f}\|_F^2}{\|f\|_F^2} \right\}, \quad (31)$$

with  $\|\cdot\|_F$  being the Frobenius matrix norm, and  $\mathcal{E}$  being the operator of expectation. In the current study, the latter is approximated by sample mean based on the results of 200 independent trials.

It has been recently argued that the NMSE may not be an optimal comparison measure as long as human visual perception is concerned. For this reason, the NMSE-based comparison has been complimented by comparing the reconstruction algorithms in terms of the SSIM index as suggested in [35].

#### D. Sparse deconvolution

Our first example is concerned with the problem of sparse deconvolution<sup>3</sup>, in which case  $\mathbf{A} \equiv \mathbf{H}$ . In this task, our main intension is to demonstrate the importance of regularization and correct modeling of noise for successful reconstruction of  $f$  in (1). Consequently, in this subsection, the performance of the PIS algorithm is compared to those of RL and GIS.

The assumption of sparsity suggests that  $f$  consists of a small number of bright sources scattered over a black background. An example of such an image is shown in the upper-right subplot of Fig. 1, where the non-zero samples of  $f$  have been generated by taking the absolute value of an i.i.d. Gaussian random variables.

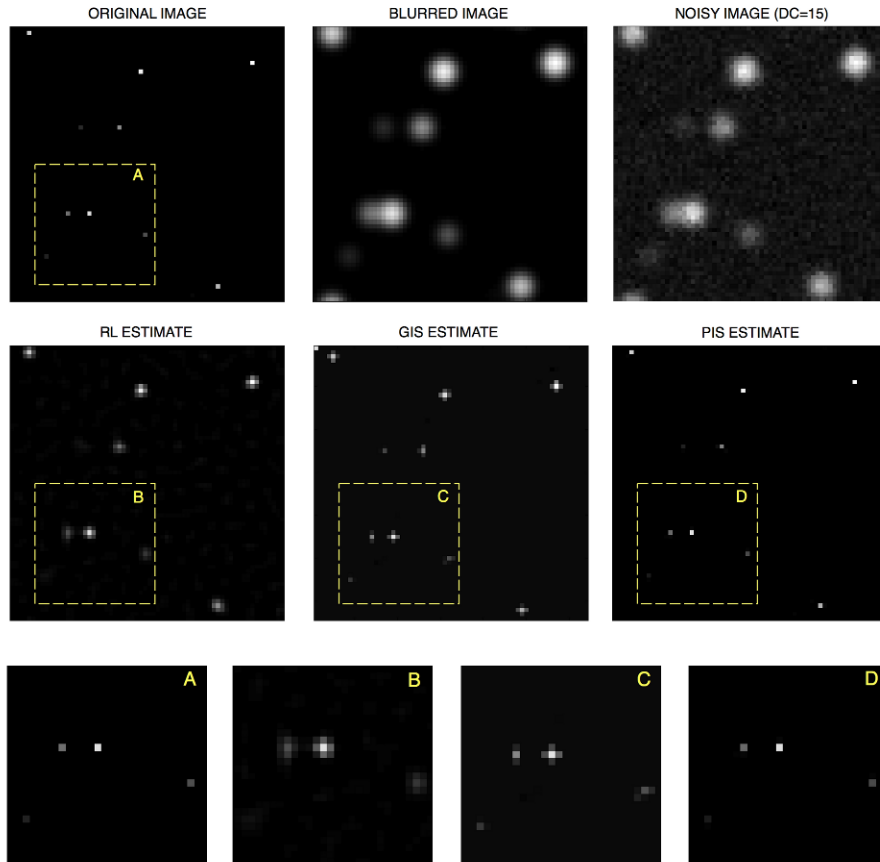


Fig. 1. (First row of subplots) Original image, blurred image, and noisy image (SNR=17); (Second row of subplots) RL reconstruction, GIS reconstruction, and PIS reconstruction; (Third row of subplots) Zoomed segments of the original and reconstructed images as indicated by the dashed rectangles.

The blurring artifact was simulated by convolving the test images with an isotropic Gaussian kernel  $h(\mathbf{x})$  whose -3 dB cut-off frequency was set to be equal to  $0.2\pi$ . As the next step, the resulting images were contaminated by

<sup>3</sup>In this case, the image  $f$  is identified with its coefficients  $c$ , i.e.  $f \equiv c$ .

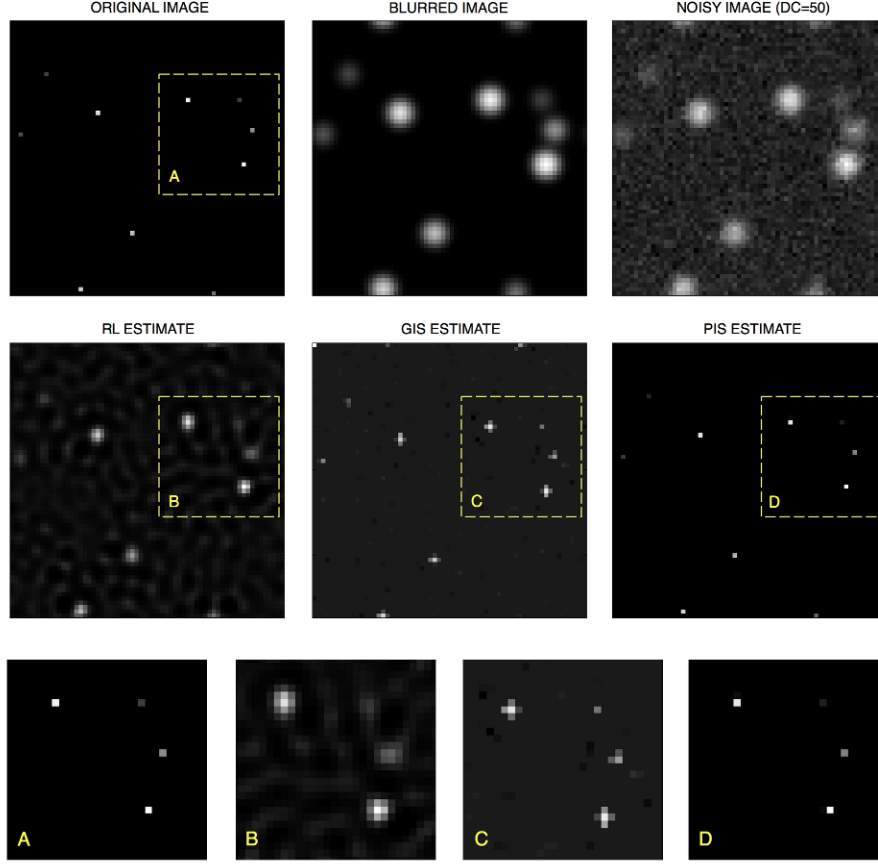


Fig. 2. (First row of subplots) Original image, blurred image, and noisy image (SNR=5.1); (Second row of subplots) RL reconstruction, GIS reconstruction, and PIS reconstruction; (Third row of subplots) Zoomed segments of the original and reconstructed images as indicated by the dashed rectangles.

Poisson noise. In this regard, it should be noted that the level of Poisson noise is defined by the corresponding value of the blurred image  $\mathbf{H}[f]$ . Since real-life images are always contaminated by background noises, the minimum value of  $\mathbf{H}[f]$  should be strictly positive. In this study, a set of three different minimum (background) values, namely 15, 30, and 50, were used. Note that since the variance of Poisson noise is equal to its mean value, higher background values will result in more severe noises. In this case, it seems to be reasonable to define the SNR as a ratio of the maximum value of  $\mathbf{H}[f]$  (i.e. 255) to its background value. According to this definition, the SNR values used in the present study were 17, 8.5, and 5.1. Two examples of simulated data images for SNR=17 and SNR=5.1 are shown in the upper-left subplots of Fig. 1 and Fig. 2, respectively.

In the case of low-pass blurs, the only possibility for  $\mathbf{H}[f]$  to generate a constant background is to require  $f(\mathbf{x}) \neq 0$  for all  $\mathbf{x} \in \Omega$ . This obviously contradicts the assumption on  $f$  to be sparse. To alleviate this deficiency, we suggest to modify the image formation model via replacing  $\mathbf{H}$  and  $f$  by  $\tilde{\mathbf{H}} \triangleq [\mathbf{H} \ 1]$  and  $\tilde{f} \triangleq [f \ f_0]^T$ , respectively, where

$f \in \mathbf{R}^{K \times L}$  is assumed to be sparse, and  $f_0 \in \mathbb{R}_+ \setminus \{0\}$  is a positive scalar defining the background level. In this case, the image formation model of (1) can be redefined as

$$g = \mathcal{P} \left\{ \tilde{\mathbf{H}}[\tilde{f}] \right\} = \mathcal{P} \left\{ \mathbf{H}[f] + f_0 \right\}. \quad (32)$$

Consequently, the reconstruction is applied with  $\tilde{\mathbf{H}}$  to recover  $\tilde{f}$ , in which case the true image is considered to be equal to the sum  $f + f_0$ .

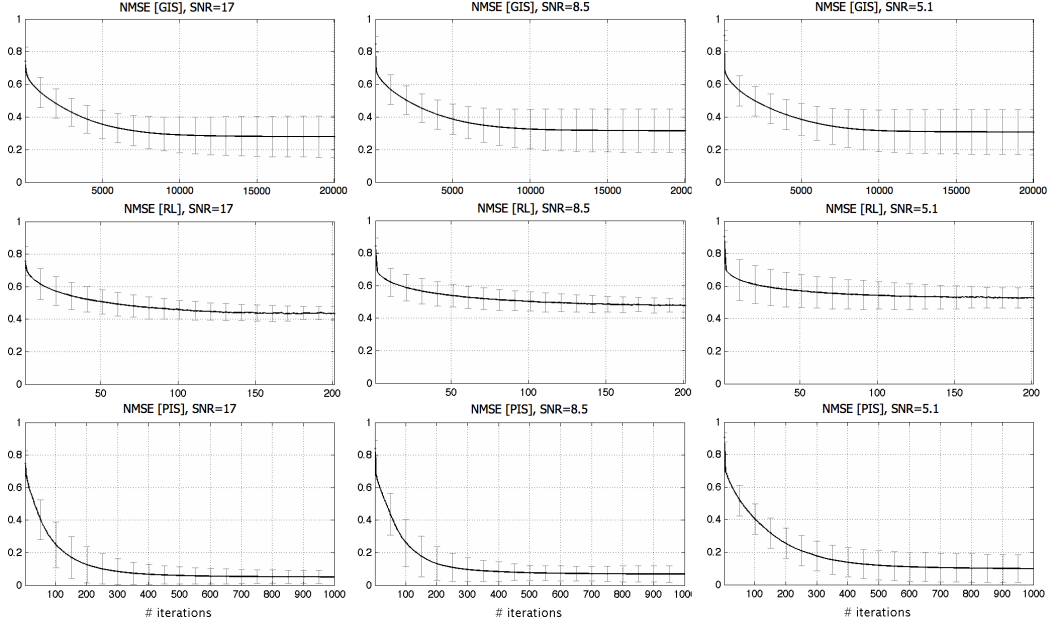


Fig. 3. (Upper row of subplots) The NMSE of GIS reconstruction as a function of the number of iterations for SNR=17, 8.5 and 5.1; (Middle row of subplots) The NMSE of RL reconstruction as a function of the number of iterations for SNR=17, 8.5 and 5.1; (Lower row of subplots) The NMSE of PIS reconstruction as a function of the number of iterations for SNR=17, 8.5 and 5.1.

The above model adjustment was applied only for the cases of GIS and PIS reconstruction, which are based on the sparsity assumption. The regularization parameter  $\gamma$  in (9) and (28) was defined to be  $1/\beta$  and  $\sigma^2/\beta$ , respectively, with  $\beta = 4.5$  and  $\sigma^2$  equal to the sample variance of the background noise. In the case of GIS, the parameter  $\mu$  was set to be equal to  $1.1 \cdot \|\tilde{\mathbf{H}}^* \tilde{\mathbf{H}}\|$ , while in the case of PIS it was chosen using Algorithm 1 with  $\alpha = 0.8$ .

Typical reconstruction results are demonstrated in Fig. 1 and Fig. 2 for SNR=17 and SNR=5.1, correspondingly. In particular, the middle row of subplots of the figures show the reconstructions obtained by the (from left to right) RL, GIS, and PIS algorithms. For the convenience of the reader, the bottom row of subplots in Fig. 1 and Fig. 2 show zoomed fragments of the original and recovered images as indicated by the dashed boxes and letters A, B, C, and D. One can see that PIS outperforms all the reference methods in terms of the resolution improvement and noise reduction.

A quantitative comparison of the reconstruction algorithms is presented in Fig. 3, which shows the NMSE as a function of the number of iterations for (from up to down) GIS, RL, and PIS, and for different values of SNR,



namely (from left to right) 17, 8.5, and 5.1. It should be noted that each value of the NMSE in Fig. 3 is a result of averaging the errors obtained in a series of independent trials, where both the true images and noises were drawn randomly.

Observing Fig. 3, one can see that PIS results in considerably lower values of the NMSE as compared to the reference methods. As well, it converges to a steady-state solution after a much smaller number of iterations as compared to the GIS algorithm (i.e. 500 vs.  $10^4$ ). Unlike the RL method, which is non-monotonely convergent in NMSE (which can not be seen in Fig. 3 since the method was terminated before the algorithm diverged), the convergence of PIS is monotone in both  $E(c)$  and NMSE.

#### E. Sparse reconstruction

In the second part of the experimental study, the PIS method was tested in application to the problem of sparse image reconstruction, where the combined operation of image synthesis and blur is represented by  $\mathbf{A} = \mathbf{H}\Phi$ . In this case, the blur model was defined by the convolution kernel  $h[i, j] = (i^2 + j^2 + 1)^{-1}$ , with  $i, j = -D, \dots, D$  and  $D \in \{2, 7\}$  [32]. The frame operator  $\Phi$  was defined to describe the translation invariant (TI) wavelet transform corresponding to the Haar wavelet. The number of wavelet resolutions was set to be equal to 4. As in the case with sparse deconvolution, the wavelet frame was extended by adding a constant vector so as to allow the image background to be modeled by a single element of the frame.

In this subsection, image reconstructions produced by the proposed and reference methods are tested using a microscopic image of glomerulus and the standard Shepp-Logan phantom, which are shown in Subplots A of Fig. 4 and Fig. 5, respectively. Similar to the case of sparse reconstruction, the images have been offset by a constant (background) value to give rise to different values of SNR. In particular, the value was adjusted to result in SNR equal to 32 (moderate noises) and 8 (strong noises). The original, blurred, and contaminated images of the glomerulus and Shepp-Logan phantom are summarized in Fig. 4 and Fig. 5 for all the tested values of  $D$  and SNR.

TABLE III  
NMSE AND SSIM VALUES OF THE RECONSTRUCTION METHODS UNDER COMPARISON USING THE SHEPP-LOGAN PHANTOM.

	L=2, SNR=32			L=2, SNR=8		
	NMSE	SSIM	NIT	MNSE	SSIM	NIT
<b>RL</b>	0.101	0.68	10	0.368	0.54	10
<b>RLTV</b>	0.141	0.72	10	0.400	0.71	10
<b>PIDsplit+</b>	0.123	0.85	5	0.872	0.68	5
<b>VSTSR</b>	0.189	0.65	300	0.377	0.65	300
<b>SPIRAL</b>	0.132	0.76	20,000	0.378	0.72	20,000
<b>PIS</b>	<b>0.09</b>	<b>0.88</b>	500	<b>0.377</b>	<b>0.84</b>	500

The reference methods used in this section were RL, RLTV, PIDsplit+, VSTSR and SPIRAL. It should be noted that the GIS method has been excluded from the current experiment, whose inappropriateness of statistical model

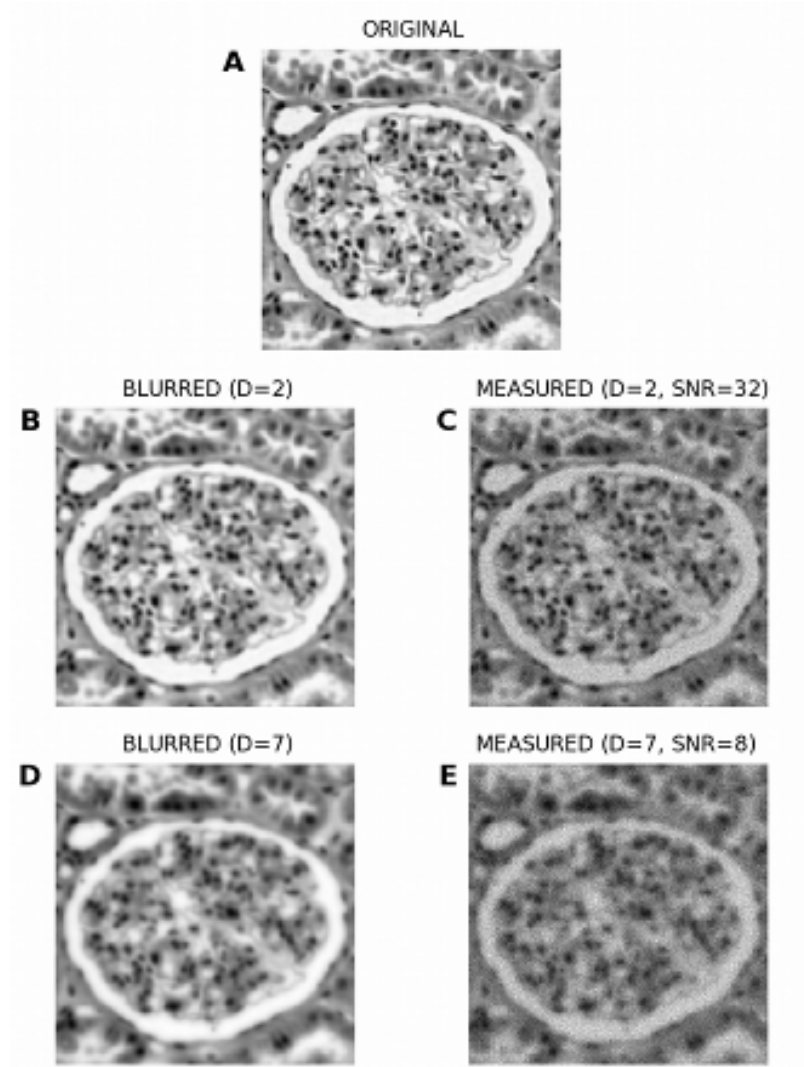


Fig. 4. (Subplot A) Original image of glomerulus; (Subplot B) Blurred image of glomerulus with  $D=2$ ; (Subplot C) Blurred and noisy image of glomerulus with  $D=2$  and  $\text{SNR}=32$ ; (Subplot D) Blurred image of glomerulus with  $D=7$ ; (Subplot E) Blurred and noisy image of glomerulus with  $D=7$  and  $\text{SNR}=8$ .

makes it a poor candidate for comparison (as demonstrated by the results of the previous section). The regularization parameters of VTSR, SPIRAL and PIS were set empirically to be equal to 0.05, 0.1 and 0.02, respectively. The regularization parameters of RLTV and PIDsplit+ were set to be equal to 0.002 and 0.01 according the guidelines provided in [14] and [17].

The SPIRAL algorithm was applied with the orthogonal Haar wavelet transform (as opposed to its stationary version used by PIS), according to the requirements specified in [21]. In this case, to alleviate the artifacts caused by the property of the orthogonal wavelet transform being translational variant, the cycle-spinning algorithm of [23] was employed, with a total number of cycles set to be equal to 20.

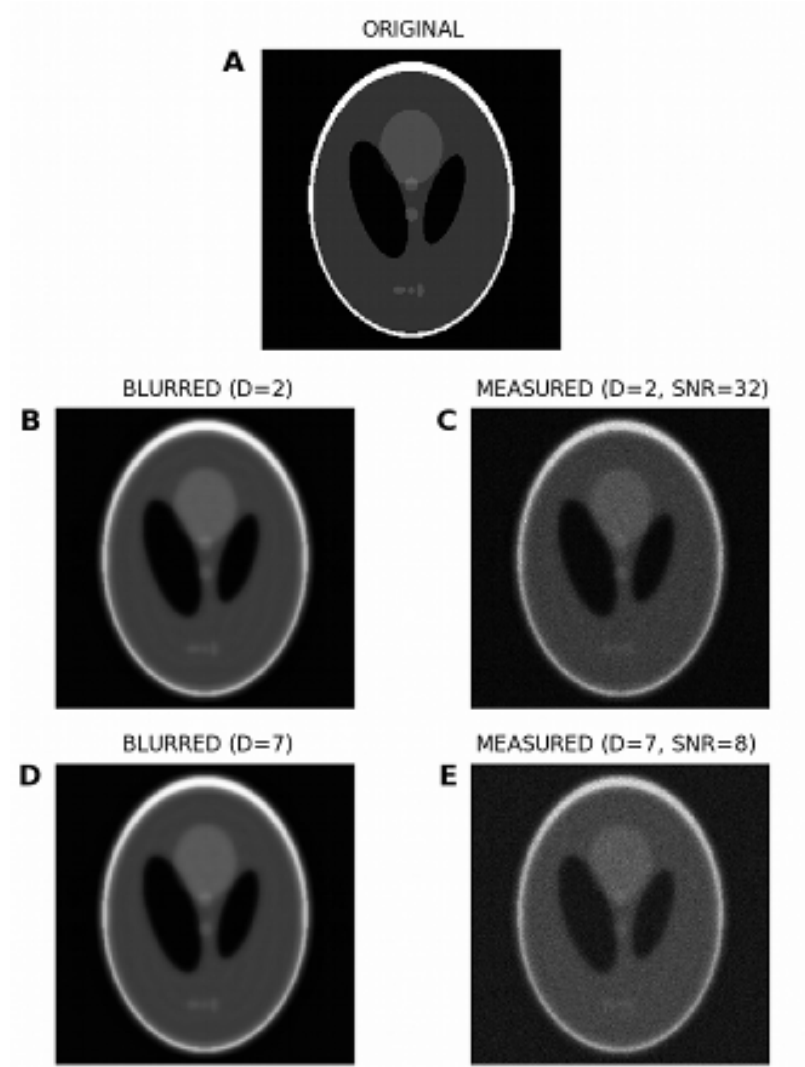


Fig. 5. (Subplot A) Original image of the Shepp-Logan phantom; (Subplot B) Blurred image of the Shepp-Logan phantom with  $D=2$ ; (Subplot C) Blurred and noisy image of the Shepp-Logan phantom with  $D=2$  and  $\text{SNR}=32$ ; (Subplot D) Blurred image of the Shepp-Logan phantom with  $D=7$ ; (Subplot E) Blurred and noisy image of the Shepp-Logan phantom with  $D=7$  and  $\text{SNR}=8$ .

In the case of the VSTSR, SPIRAL and PIS algorithms, their execution was terminated automatically at the point when the relative change  $\|f_{t+1} - f_t\|_F / \|f_t\|_F$  between iterations  $t$  and  $t+1$  was observed to drop below a threshold of  $10^{-6}$ . Unfortunately, the same stopping criterion could not be applied to the RL, RLTV, and PIDsplit+ methods, whose steady-state estimation was found to be unacceptably noisy. For this reason, these algorithms were terminated earlier, at the point when their corresponding NMSE reached their minimum values. It should be noted that, since the computation of NMSE requires the knowledge of a true image, the “NMSE-optimal” convergence cannot be considered as a practical tool. Therefore, the results of RL, RLTV, and PIDsplit+ methods reported in this section may not be reproduced in a real-life scenario.

TABLE IV  
NMSE AND SSIM VALUES OF THE RECONSTRUCTION METHODS UNDER COMPARISON USING A GLOMERULUS IMAGE; NMSE VALUES  
APPEAR AFTER MULTIPLICATION BY  $10^2$ .

	L=4, SNR=32			L=7, SNR=8		
	NMSE	SSIM	NIT	MNSE	SSIM	NIT
<b>RL</b>	0.280	0.86	10	0.772	0.77	10
<b>RLTV</b>	0.275	0.87	10	0.755	0.78	10
<b>PIDsplit+</b>	42.54	0.90	4	47.51	0.78	4
<b>VTSR</b>	9.150	0.78	300	9.593	0.62	300
<b>SPIRAL</b>	0.303	0.86	50,000	0.737	0.81	50,000
<b>PIS</b>	<b>0.210</b>	<b>0.92</b>	400	<b>0.729</b>	<b>0.86</b>	400

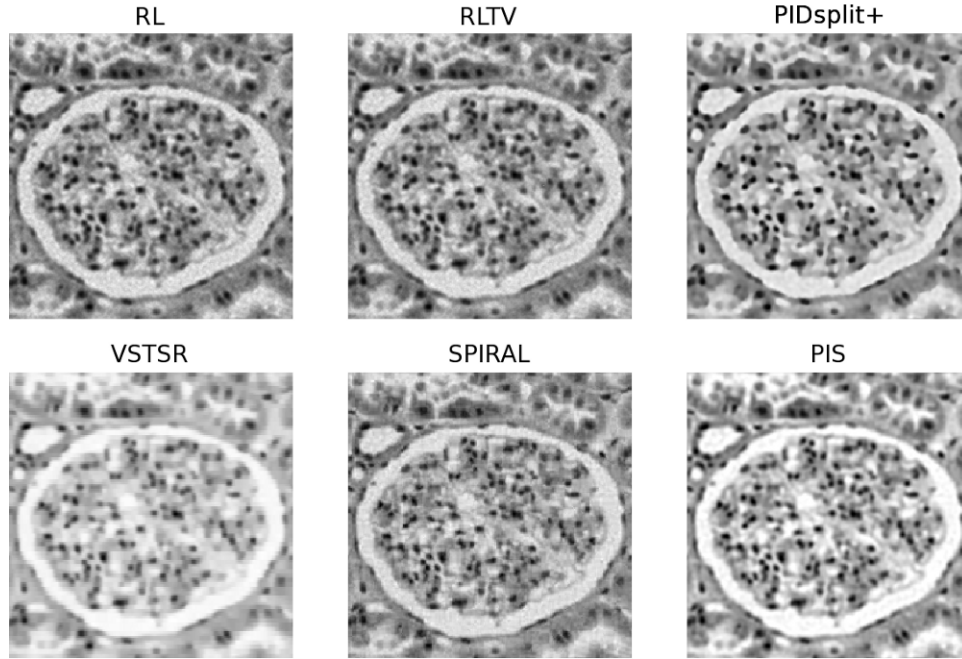


Fig. 6. Image reconstruction results corresponding to Fig.4 with  $D=2$  and  $SNR=32$ . (Upper row of subplots) RL, RLTV and PIDsplit+ estimates; (Lower row of subplots) VTSR, SPIRAL and PIS estimates.

For the case of glomerulus, the reconstructions obtained with the proposed and reference methods are summarized in Fig. 6 (for  $D = 2$ ,  $SNR=32$ ) and Fig.7 (for  $D = 7$ ,  $SNR=8$ ). Moreover, Fig. 8 and Fig. 9 depict the reconstructions of the Shepp-Logan phantom for the cases of  $D = 2$ ,  $SNR=32$  and  $D = 7$ ,  $SNR=8$ , respectively. Analyzing these results, one can clearly see that, in all the above cases, the PIS algorithm yields reconstructions of superior quality (in terms of the resolution and contrast gain), as compared to the reference methods. This observation is further supported by the quantitative measures of Tables III and IV, which compare the estimation results in terms of the NMSE, SSIM index, and the number of iterations. As evidenced by the tables, the PIS method produces the

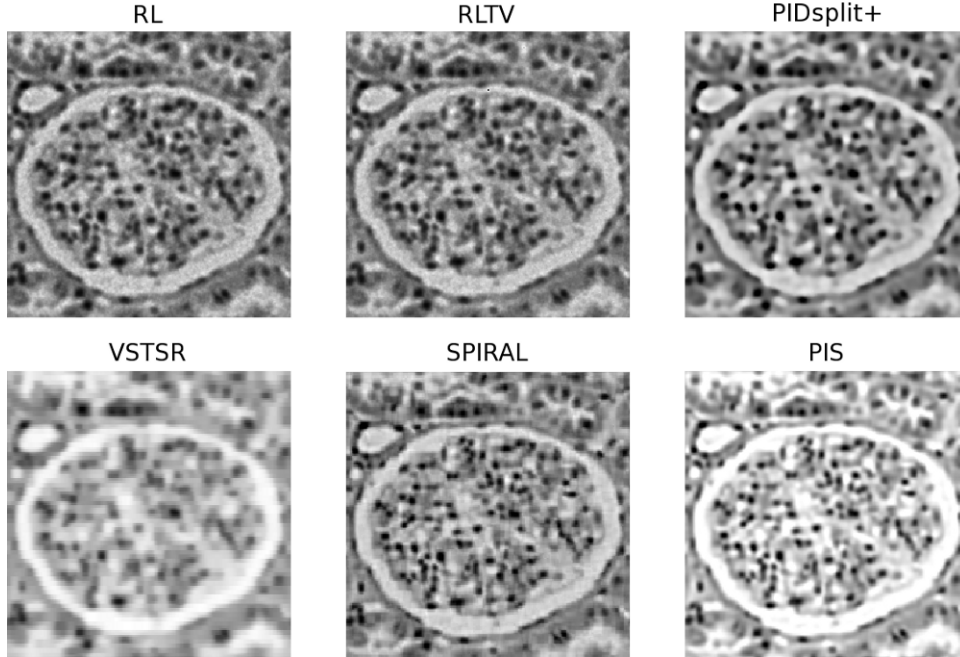


Fig. 7. Image reconstruction results corresponding to Fig.4 with  $D=7$  and  $SNR=8$ . (Upper row of subplots) RL, RLTV and PIDsplit+ estimates; (Lower row of subplots) VSTSR, SPIRAL and PIS estimates.

lowest NMSE and the largest SSIM index among all the methods under comparison. As to the number of iterations required by PIS, one can see (with a reference to Table II) that the method has a computational complexity either comparable or lower than that of the reference methods.

## VI. DISCUSSION AND CONCLUSIONS

In the current paper, a new approach to the problem of denoising/reconstruction of digital images was presented. The method has been derived based on the framework of MAP estimation under the assumption of Poisson noise contamination. Such noise models are known to be standard in many important image modalities, including optical, microscopic, turbulent, and nuclear imaging, just to name a few. Moreover, whilst many of the existing solutions to the problem of enhancement of Poissonian images take advantage of certain simplifying assumptions about the noise nature, the proposed technique is optimized to deal with the realistic noise model at hand.

Another advantage of the proposed method consists in the generality of its formulation. The latter allows applying the same reconstruction procedure to a number of different settings, such as image denoising or image enhancement through deconvolution. Furthermore, the prior assumptions made by the method regarding the nature of recovered images are general as well. Specifically, the images are assumed to admit a sparse representation in the domain of a properly chosen linear transform. Note that the reasonability of the above *a priori* modeling is firmly supported by the recent advances in the theory of sparse representation.

Yet another critical advantage of the proposed PIS algorithm is in its algorithmic structure, which exploits the

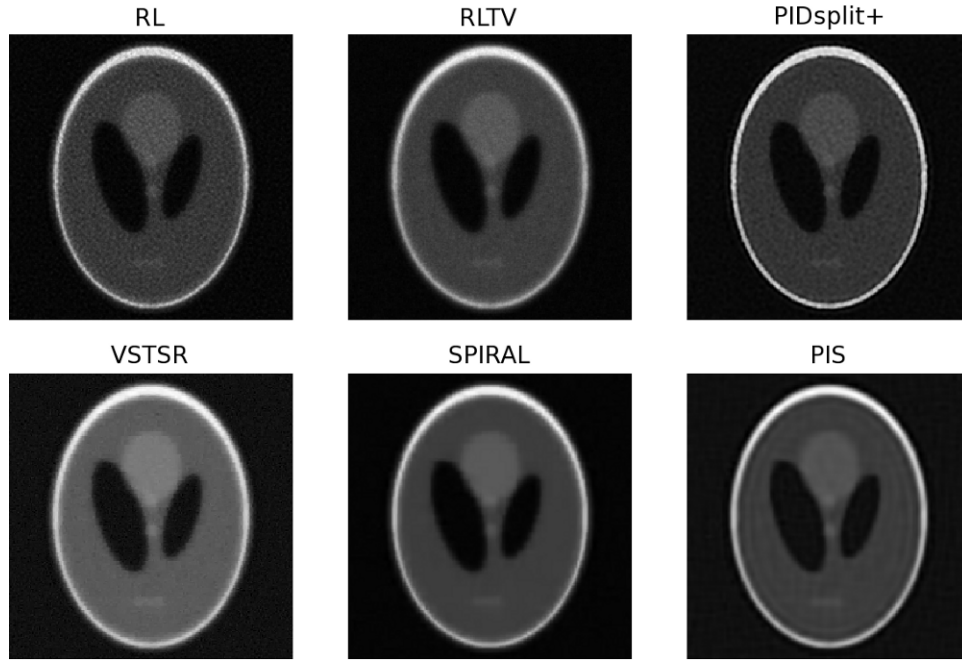


Fig. 8. Image reconstruction results corresponding to Fig.5 with  $D=2$  and  $SNR=32$ . (Upper row of subplots) RL, RLTV and PIDsplit+ estimates; (Lower row of subplots) VSTSR, SPIRAL and PIS estimates.

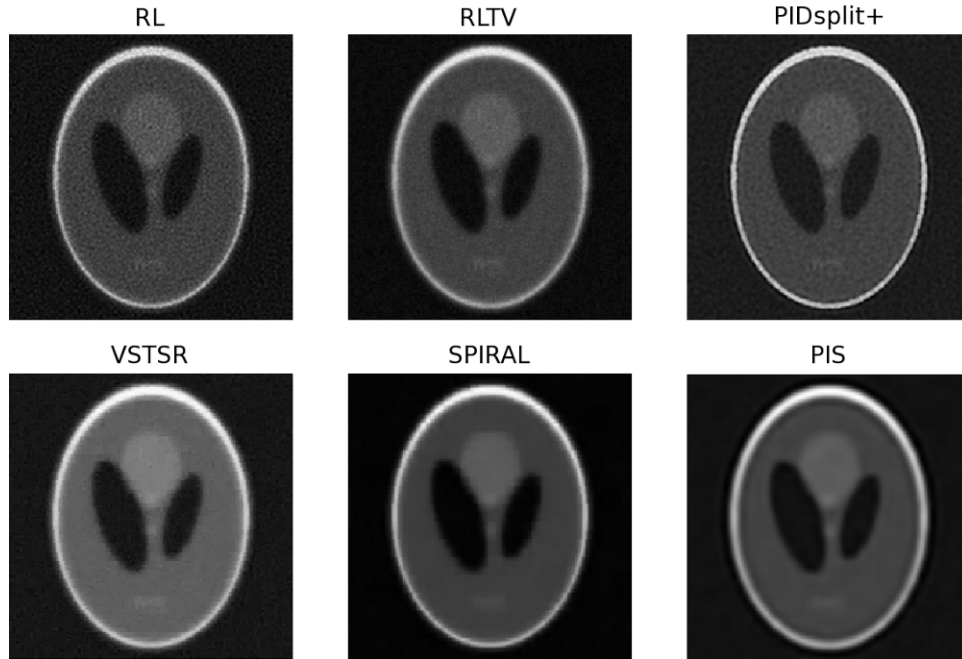


Fig. 9. Image reconstruction results corresponding to Fig.5 with  $D=7$  and  $SNR=8$ . (Upper row of subplots) RL, RLTV and PIDsplit+ estimates; (Lower row of subplots) VSTSR, SPIRAL and PIS estimates.

idea of iterative shrinkage. The latter allows solving non-smooth optimization problems at the computational cost of a steepest descent procedure. Consequently, the computational load required by the proposed method is relatively small, which allows the method to be applied for the solution of large-scale problems and/or for processing of large data sets.

It was shown conceptually and experimentally that the performance of the PIS algorithm is superior to a number of alternative approaches. A series of comparison tests have been performed, in which PIS was shown to outperform the reference methods in terms of both NMSE and SSIM index measures. Moreover, as opposed to the alternative methods, the PIS algorithm has always been capable of converging in a stable and robust manner to a useful reconstruction result.

We believe that the convergence speed of PIS iterations can be further improved via employing the line search strategy as detailed in [56]. It seems also to be possible to increase the accuracy of PIS estimation by using more sparsifying multiresolution transforms as compared to separable wavelets [57]. Finally, we note that the sparseness of representation coefficients appears to be a rather weak constraint to be used in the case of poorly conditioned convolution operators  $\mathbf{H}$ . This problem, therefore, should be resolved by using different prior models, whose formulation constitutes one of the directions of our current research.

## REFERENCES

- [1] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 669–680, Dec. 1995.
- [2] M. Yavuz and J. A. Fessler, "Statistical image reconstruction methods for randomsprecattered PET scans," *Med. Im. Anal.*, vol. 2, no. 4, pp. 369–378, 1998.
- [3] H. H. Bauschke, D. Noll, A. Celler, and J. M. Borwein, "An EM algorithm for dynamic SPECT," *IEEE Trans. Med. Imag.*, vol. 18, no. 3, pp. 252–261, Mar. 1999.
- [4] R. A. Boie and I. J. Cox, "An analysis of camera noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 6, pp. 671–674, Jun. 1992.
- [5] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 3, pp. 267–276, Mar. 1994.
- [6] T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: Maximum-likelihood approach," *J. Opt. Soc. Am. A*, vol. 9, no. 7, pp. 1052–1061, Jul. 1992.
- [7] H. Bradt, *Astronomy methods: A physical approach to astronomical observations*. Cambridge University Press, 2004.
- [8] M. C. Roggemann and B. Welsh, *Imaging through turbulence*. CRC Press, 1996.
- [9] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC Press, 1998.
- [10] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am. A*, vol. 62, no. 1, pp. 55–59, 1972.
- [11] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, no. 6, pp. 745–754, 1974.
- [12] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, Oct. 1982.
- [13] M. Mignotte, J. Meunier, J.-P. Soucy, and C. Janicki, "Comparison of deconvolution techniques using a distribution mixture parameter estimation: Application in single photon emission computed tomography imagery," *J. Electron. Imaging*, vol. 11, no. 1, pp. 11–24, Jan. 2002.
- [14] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Richardson-Lucy algorithm with total variation regularization for 3-D confocal microscope deconvolution," *Miscosc. Res. Techniq.*, vol. 69, no. 4, pp. 260–266, Apr. 2006.

- [15] M. Figueiredo and J. Bioucas-Dias, “Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization,” in *IEEE Workshop on Statistical Signal Processing - SSP2009*, Cardiff, Wales, UK, 2009.
- [16] T. Goldstein and S. Osher, “The split Bregman method for L1 regularized problems,” *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
- [17] S. Setzer, G. Steidl, and T. Teuber, “Deblurring Poissonian images by split Bregman techniques,” in *Preprint, University of Mannheim*, 2009.
- [18] D. Lingenfelter, J. Fessler, and Z. He, “Sparsity regularization for image reconstruction with Poisson data,” in *Proceedings of SPIE (Computational Imaging VII)*, San Jose, California, USA, 2009.
- [19] F. X. Dupe, M. J. Fadili, and J. L. Strach, “Image deconvolution under Poisson noise using sparse representations and proximal thresholding iteration,” in *ICASSP 2008*, Las-Vegas NV, USA, Apr. 2008.
- [20] F. J. Anscombe, “The transformation of Poisson, binomial, and negative binomial data,” *Biometrika*, vol. 35, pp. 246–254, 1948.
- [21] Z. Harmany, R. Marcia, and R. Willett, “Sparse Poisson intensity reconstruction algorithms,” in *IEEE Stat. Sig. Proc. Workshop*, Cardiff, Wales, UK, Aug. 2009.
- [22] R. Willet and R. D. Nowak, “Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging,” *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 332–350, 2003.
- [23] R. Coifman and D. L. Donoho, *Translation invariant denoising*, ser. Wavelets and statistics. New York, USA: Springer-Verlag, 1995.
- [24] P. Fryzlewicz and G. P. Nason, “A Haar-Fisz algorithm for Poisson intensity estimation,” *J. Comput. Graph. Statist.*, vol. 13, pp. 621–638, 2004.
- [25] M. Jansen, “Multiscale Poisson data smoothing,” *J. Roy. Statist. Soc. B*, vol. 68, no. 1, pp. 27–48, 2006.
- [26] P. Fryzlewicz, V. Delouille, and G. P. Nason, “GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform,” *J. Roy. Statist. Soc. C*, vol. 56, no. 1, pp. 99–116, 2007.
- [27] B. Zhang, J. M. Fadili, and J.-L. Starck, “Wavelets, ridgelets, curvelets for Poisson,” *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1093–1108, Jul. 2008.
- [28] R. D. Nowak and R. G. Baraniuk, “Wavelet-domain filtering for photon imaging systems,” *IEEE Trans. Image Processing*, vol. 8, no. 5, pp. 666–678, May 1999.
- [29] E. D. Kolaczyk, “Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics,” *Astrophys. J.*, vol. 534, pp. 490–505, 2000.
- [30] K. E. Timmermann and R. D. Nowak, “Multiscale modeling and estimation of Poisson processes with application to photo-limited imaging,” *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 846–862, Apr. 1999.
- [31] R. D. Nowak and E. D. Kolaczyk, “A statistical multiscale framework for Poisson inverse problems,” *IEEE Trans. Inform. Theory*, vol. 46, no. 5, pp. 1811–1825, Aug. 2000.
- [32] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, “A wide-angle view at iterated shrinkage algorithms,” in *Proceedings of SPIE (Wavelet XII)*, San-Diego CA, USA, 2007.
- [33] I. Daubechies, M. Defrise, and C. DeMol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *arXiv:math/0307152v2*, 2003.
- [34] T. Blumensath and M. E. Davies, “Iterative thresholding for sparse approximations,” *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, Dec. 2008.
- [35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [36] O. Christensen, *Frames and bases: An introductory course*. Birkhauser, 2008.
- [37] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by Basis Pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, Mar. 2001.
- [38] D. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [39] E. Candes and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies,” *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [40] P. Bofill and M. Zibulevsky, “Underdetermined blind source separation using sparse representations,” *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, Nov. 2001.



- [41] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Networks*, vol. 16, no. 4, pp. 992–996, Jul. 2005.
- [42] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Med. Imag.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [43] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [44] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [45] J. L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Adv. Imag. Electron. Phys.*, vol. 132, 2004.
- [46] O. Michailovich and D. Adam, "A high-resolution technique for ultrasound harmonic imaging using sparse representations in Gabor frames," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1490–1503, Dec. 2002.
- [47] B. P. Carlin and T. A. Louis, *Bayes and empirical Bayes methods for data analysis*, ser. Monographs on Statistics and Applied Probability. Chapman and Hall, 1996, vol. 69.
- [48] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.
- [49] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [50] D. Bertsekas, *Nonlinear programming*. Belmont, MA: Athena Scientific, 1999.
- [51] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.
- [52] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proceedings of ICIP*, Genova, Italy, 2005.
- [53] R. C. Gonzales and R. E. Woods, *Digital image processing*. Prentice hall, 2002.
- [54] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.
- [55] E. J. Candes and D. L. Donoho, "Ridgelets: The key to high-dimensional intermittency?" *Philos. Trans. R. Soc. London, Ser. A*, vol. 357, pp. 2495–2509, 1999.
- [56] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.
- [57] E. J. Candes and F. Guo, "New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction," *Signal Processing*, vol. 82, pp. 1519–1543, Mar. 2002.